**Trust Your Science? Open your Data and Code:**

Speaking of reproducibility, it is a common practice nowadays specifically not to copy things but to learn things in a different way in my view, just because you are doing something again which has been done already doesn’t mean it is being copied, it means that work is being done in a different environment with different ideology but with the same logic. For instance, consider a scenario of making a Sedan, and how they have evolved from the older times, earlier they just had the engines, window operating switches, and a few more limited features, but if you see a modern Sedan these days they have a lot of features in it like automatic window switches, Sunroofs, you can make a call just by clicking on one button in your dashboard, also look at Tesla they even automated the driving, just adjust your car onto autopilot driving mode and the car is going to take you to your destination without you bothering to drive. This is a perfect example of reproducibility; you are not copying but developing and trying to improve things that have already been there around you.

Similarly, the first 5 paragraphs are elaborating on what reproducibility is and which branches of Science are trying to embrace it. There is one paragraph that is talking about how hard it is sometimes to publish the work without errors, well in those cases verifying and having discussions on what has been done could help in overcoming all the problems.

The next part Reproducibility, Replicability, and Repeatability are about, making your work public so that it can be reused to accomplish outstanding achievements. In the Author’s case, they released their MATLAB codes and made them public so that they can be regenerate the figures, doing that will help the process of the work in an efficient way because third parties working on the existing work will allow them to think in a different way and help to get additional points on how to accomplish the work in a unique way. Again, as we can see that the last paragraph is concentrating on how the reproduced work is thought to be accurate just because one is redoing it, then there is a sentence that talks about verifying validating, and performing error qualification on the developed work, which in my view is not a bad idea to consider. Doing that will allow one to get more input from other peers working on the same thing but in a different way.

After that, the data evaluation standards section talks about the data related to Genome decoding being shared amongst the community that was gathered in Bermuda and doing the same thing for the next three times but in a safe and legal way. This resulted in understanding the data, making updates to the dataset, and performing the analysis on the available data, etc., for effective reproducibility. In the last few paragraphs, the discussion is that making the data open for reproducibility is not a convenient way to concentrate because analyzing the research data needs deeper justification and understanding of course, but what I can say is that reproducibility should be done only in the areas where it is efficient to do so, an example that I already mentioned in the first paragraph. One more point I would like to add is that it can also be used as a learning process by those who are trying to break into their respective domains.

**What is a Statistical Project?**

Basically, the first paragraph is about what a Statistical Project is, it is a report which explains a research question using statistical techniques to present the findings. Then in the first few paragraphs of Data-Based Problem-solving, the discussion is on how the students should think of the questions to address a research question. Initially, it will be a bit complicated task but eventually, it will be the right approach to follow. If the question is decided, the student should be keen enough to understand the source of the dataset and analyze it. After finalizing the data, the preprocessing and understanding of the key features in the dataset is the student’s responsibility to avoid the circumstances after going further into the project. In my view, I believe what’s being said as there is no use in working on a project without knowing the main components in the dataset and it will not help you to get to a conclusion.

One should have strong graphical representation skills to analyze the data and clarification what model or approach to use for the research question. The main point to focus is on the representation of the research problem in an accurate way and not deviate from the idea.

Then comes the written report that must consist of the problem that you are working on, show the research that you performed, details on where the data was collected and stored, after all the analysis getting to the conclusion of the problem statement and satisfying it.

As stated in the final note, schools should involve statistical activities from an early age. This is a good idea because statistical knowledge will only enhance the logical thinking in children at a younger age, and by the time they grow, they will already have the knowledge to deal with statistics and work on projects independently. It will also improve the student's communication skills.

**Where does a Statistician Fir in the Big Data Era?**

This article discusses whether statisticians can be a part of Big Data or not. The first three paragraphs explain what will be discussed in the rest of the article and how statistics have previously been a part of the science of data, which is now known as Data Science. I believe statistics is a subset of data science because a lot of data analysis involves aggregation, predictive analysis, and so on.

The author then continued to explain the four skills that a statistician scientist should possess. The first is Numeracy, wherein he believes that mathematics and numerical computation are areas in which a statistician should concentrate to tackle real-world problems. Because most data is usually messy, statisticians must develop various tools and procedures to make better predictions. From this, we can determine that whatever models are used to complete projects in Data Science/ Machine Learning/Deep Learning, such as Trees, Regression Models, Classification Models, and so on, are all ideas developed by statisticians in the past.

Moving on to Articulacy, where is it said that this skill is required not only for statisticians but also everyone in the working environment. Clean and efficient communication is the key to articulacy because making your work public isn’t the only sufficient way to make everyone understand it you should be able to publish it in a way that even a non-technical guy can get an idea without you telling him just looking at your work. Whilst working on the statistical model he should be able to understand discipline and the questions of other scientists too. One more skill in articulacy is engaging presentation, well in my view this is the main thing a statistician should focus on as this is the one and only efficient and clear way to preset your work. Overall, a so-called synonym that I can relate to articulacy is communication.

Following that is literacy, which means reading and writing in the sense if there is an unavoidable situation, statisticians will have to read all the available work on that issue and be able to write a solution to it mostly code. Essentially after the formulation of a method, he should be able to build a model and perform computations on it. Statisticians often work with already clean data.

The fourth skill that the author has mentioned is Graphicacy usually visualizing the available data according to the author, there is a distinction between statistical visualization and data visualization. One example of this that I can give is calculating the mean value of the target variable and trying to visualize it like the average price of households in specific areas. Another statistical plot that I get in mind is the correlation heat map that gives the relationship between the available numerical features in the dataset.

**Feedback:**

My understanding from this article is that statisticians do have a role in Data and using statistical methods in the projects will help discover new findings.

Reproducibility must be valued in such a way that if it is used to learn, it is fine; however, if it is being used to copy instead of learning, it is inappropriate.